

REMARKS

Applicant respectfully requests further examination and reconsideration in view of the remarks set forth below. Claims 1-60 were previously pending in this application. In the Office Action, Claims 1-5, 7-9, 12, 17, 18, 20, 27, 32-34, 37-39, 45, and 58-60 are rejected, and Claims 6, 16, 19, 22-26, 28, 41, 44, 47-51, 53, 54, 56, and 57 have been withdrawn in accordance with the previously filed restriction requirement. Also within the Office Action, Claims 52 and 55 are allowed, and Claims 10, 11, 13-15, 21, 29-31, 35, 36, 40, 42, 43, and 46 are objected to and would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims. Each of the rejections is fully addressed below. Accordingly, Claims 1-5, 7-15, 17-18, 20-21, 27, 29-40, 42-43, 45-46, 52, 55, and 58-60 are now pending in this application.

The present invention is a system and method for speech recognition using an adaptive multi-pass technique. The system includes an input device coupled to a source of spoken input for receiving the spoken input. A processor coupled to the input device performs a first pass speech recognition technique on the spoken input and forms first pass results. The first pass results can include a number of alternative speech expressions, each having an assigned score representative of the certainty that the corresponding expression correctly matches the spoken input. In the preferred embodiment, scores for alternative expressions and differences between such scores are utilized to determine whether to perform another speech recognition pass.

Preferably, the first pass is performed by a simpler speech recognition technique which narrows the possibilities for expressions which match the spoken input, while the second pass is performed only when necessary and by a more complex speech recognition technique which operates on only the narrowed possibilities. Both first and second pass are used to recognize the same language and the same vocabularies. However, the second pass uses more complex

techniques.

Hedin teaches a system for enabling low power terminals to access and control remote server applications via a voice controlled interface. A Terminal Part (TP) 203 and Terminal Application Part (TAP) 201 embody a client terminal (client part 101), which is coupled to a Remote Application Part (RAP) 205. The TP 203 receives speech from a user. The input speech is provided to the TAP 201, where each word within the input speech is isolated (Hedin, col. 7, lines 7-11). Each "isolated word" is supplied to an automatic speech recognition system (ASR) 227 within TAP 201 for isolated word recognition analysis (Hedin, col. 7, lines 14-16).

If the isolated word can not be recognized by the ASR 227, then the audio encoded data corresponding to the unrecognized isolated word is packaged and sent to the RAP 205, which includes a different speech recognition system, ASR 307. In order to pass the unrecognized isolated word on to the RAP 205, the audio encoded data from the start/stop detector and recording unit 225 is formatted as MIME types by a MIME formatting unit 247 in the TAP 201 (Hedin, col. 8, lines 30-33). The ASR 307 then attempts to recognize the audio encoded word that was not recognized in the TAP 201, that is, words that were transferred to the RAP 205 as MIME types (Hedin, col. 9, lines 12-16).

In summary, Hedin teaches a first ASR (ASR 227 within TAP 201) that determines if a received spoken word is a match or is not a match to a limited vocabulary. Any words that are not recognized by the first ASR are repackaged as MIME-formatted audio encoded data, which are sent to a second ASR which tries to recognize a different vocabulary (ASR 307 within RAP 205). MIME-formatted audio encoded data are merely audio signals. The second ASR receives the audio signal as a blank slate. That is, there are no previous analysis of the unrecognized words (otherwise referred to as preliminary matching speech expressions) accompanying the audio signal. The second ASR receives the audio signal as if the second ASR is receiving the audio signal directly from the audio signal source. The second ASR then proceeds to analyze the received audio signal. Hedin does not teach that when there is not a definitive match made by the first ASR that one or more "possible matches" are made, e.g. first pass results, and that these

possible matches are forwarded to the second ASR, where the second ASR attempts to recognize the possible matches (first pass results). Instead, Hedin teaches that the audio encoded data corresponding to an unrecognized word is sent to the second ASR. The audio encoded data is the actual audio representation of the originally spoken word. The audio spoken data is not a possible recognized match output by the first ASR. Thus, the first and the second passes are used to recognize different vocabularies from each other.

Rejections under 35 U.S.C. § 102

Within the Office Action, Claims 58-60 stand rejected under 35 U.S.C. 102(e) as being anticipated by U.S. Patent No. 6,185,535 issued to Hedin et al. (hereafter "Hedin"). The Applicant respectfully traverses this rejection.

Hedin teaches that an isolated word is passed to the first ASR 227. The ASR 227 includes a feature vector extraction unit 229 which receives the isolated word and maps it into a vector space that is suitable for use by a feature matching and decision unit 231 (Hedin, col. 7, lines 16-20; Figure 2). The feature matching and decision unit 231 compares the feature vector supplied at the output of the feature vector extraction unit 229 with feature vectors supplied by the TAP reference database 233 (Hedin, col. 7, lines 33-36). There is no hint, teaching, or suggestion within Hedin to indicate that the ASR 227 selects a speech recognition technique based upon a result of a speech recognition technique performed on prior spoken input. Rather, Hedin uses the same recognition technique regardless of the grammar or vocabulary being changed as a result of the second pass needing to be performed. The ASR 227 performs the same speech recognition process for all input speech.

Within the Office Action, it is stated on page 6, paragraph number 7, that Hedin teaches "the selection of a speech recognition pass based on previous results" and that this is the same as the claimed limitation of selectively performing a first pass speech recognition technique on the spoken input based upon a result of a speech recognition technique performed on prior spoken input from the source. To support this assertion, it is further stated within the Office Action that

Hedin teaches when the first pass speech recognition fails, the user is offered to input the word in a different manner, such as repeating the word, spelling the word, or typing the word which would require a particular means of recognition, either utterance recognition, spelling recognition, or text recognition. It is therefore concluded that the type of recognition is dictated by the type of input from the user after an initial recognition. The Applicant contends that this is not the same as the aforementioned claimed limitation. The independent Claims 58 and 60 are each directed to selectively performing a speech recognition technique. Spelling recognition and text recognition, as suggested, are clearly not speech recognition. Therefore, as related to speech recognition, Hedin teaches the same speech recognition technique for all input speech. There is no selective performing of speech recognition techniques based upon prior spoken input.

Further, the independent Claim 59 includes the limitation of selectively performing a first pass speech recognition technique on the spoken input based upon information obtained regarding a speaker of the spoken input. There is no hint, teaching, or suggestion within Hedin that information is obtained regarding the speaker, and that the obtained information is used to selectively perform a speech recognition technique. Further, there is no indication within the Office Action that Hedin teaches such a limitation.

The independent Claim 58 includes a method of recognizing spoken input received from a source of the spoken input. The method comprises receiving the spoken input from the source of the spoken input, selectively performing a first pass speech recognition technique on the spoken input based upon a result of a speech recognition technique performed on prior spoken input from the source, and performing a second pass speech recognition technique on the spoken input. As discussed above, Hedin teaches a single speech recognition technique, which can not be modified and is not based on the speech recognition results from a prior spoken input. For at least these reasons, the independent Claim 58 is allowable over of Hedin.

The independent Claim 59 includes a method of recognizing spoken input received from a source of the spoken input. The method comprises receiving the spoken input from the source of the spoken input, selectively performing a first pass speech recognition technique on the

spoken input based upon information obtained regarding a speaker of the spoken input, and performing a second pass speech recognition technique on the spoken input. As discussed above, Hedin teaches a single speech recognition technique, which can not be modified and is not based on information obtained regarding the speaker of the spoken input. For at least these reasons, the independent Claim 59 is allowable over of Hedin.

The independent Claim 60 includes a method of recognizing spoken input received from a source of the spoken input. The method comprises receiving the spoken input from the source of the spoken input, selectively modifying a first pass speech recognition technique to be performed on the spoken input based upon a result of a speech recognition technique performed on prior spoken input from the source, and performing the first pass speech recognition technique on the spoken input. As discussed above, Hedin teaches a single speech recognition technique, which can not be modified and is not based on the speech recognition results from a prior spoken input. For at least these reasons, the independent Claim 60 is allowable over of Hedin.

Rejections under 35 U.S.C. § 103

Within the Office Action, Claims 1-5, 7-9, 12, 17, 18, 20, 27, 32-34, 37-39, and 45 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Hedin in view of U.S. Patent No. 5,526,463 issued to Gillick et al. (hereafter "Gillick"). The Applicant respectfully traverses this rejection.

Within the Office Action, it is acknowledged that Hedin does not teach limiting a first pass speech recognition to a subset of matches. However, it is suggested that a fast match, or course, recognition technique of Gillick can be used to limit the initial results of Hedin to a shortened list of possible candidates before using the second ASR (ASR 307 within RAP 205) of Hedin. The Applicant contends that such a combination does not result in a properly functioning system, and as such, is not a proper combination.

As discussed above, Hedin teaches receiving input speech from a user, the input speech is segmented into isolated words, and an attempt to recognize each isolated word is made by a first

ASR. Any words that are not recognized by the first ASR are sent as an audio signal to a second ASR. The second ASR then attempts to recognize the sent audio signal. No preliminary matches are sent with the audio signal to the second ASR. The second ASR is configured to recognize the received audio signal as if received directly from an audio source. Adding the output from the first recognizer is not done because the second recognizer uses a different vocabulary.

Gillick teaches a fast match system in which a first, or course, speech recognition technique is performed on input speech to provide a reduced set of possible matches. This course search is not configured to find a single accurate match, but instead to narrow the match possibilities. The set of possible matches is then sent to a second, or fine, speech recognition system and a more detailed speech recognition technique is performed.

The proposed combination necessitates the use of the Gillick first, or course, speech recognizer to perform a first pass. However, as discussed above, a first pass within Gillick does not result in an accurate, singular match. Using the first speech recognizer of Gillick requires a second pass to achieve an accurate, singular result. In the proposed combination, the second pass is performed by the second ASR of Hedin. However, the second ASR of Hedin is configured to receive an unknown, previously unrecognized, audio file. The second ASR is to use its entire vocabulary to identify the received unknown audio file. Sending a set of possible matches, as generated by the first speech recognizer of Gillick, does not make sense as applied to the expected received input by the second ASR of Hedin. The required input for the second ASR of Hedin is an unknown audio file, whereas the input provided by the first speech recognizer of Gillick is a list of possible matches, or possible known matches, albeit "known" with a questionable degree of accuracy. The expected input and the provided input are not the same. Therefore, the combination of Hedin in view of Gillick would not function as proposed, and as such is not a proper combination.

There is no hint, teaching, or suggestion within Hedin as to how a received input of the type generated by the first speech recognizer of Gillick is to be processed. Further, since Hedin

teaches that the received audio input is MIME-formatted audio encoded data, it is not clear how the list of possible matches from Gillick is to be formatted and sent to the second ASR of Hedin, and what content the list of possible matches would include in order to be used in the speech recognizing process by the second ASR.

The independent Claim 1 includes a speech recognition system for recognizing spoken input received from a source of the spoken input coupled to the speech recognition system. The speech recognition system comprises input means for receiving the spoken input from the source of the spoken input, and processing means coupled to the input means for performing a first pass speech recognition technique on the spoken input and for forming first pass results, wherein the first pass results define one or more preliminary matching speech expressions, further wherein the processing means selectively performs a second pass speech recognition technique on the spoken input according to the first pass results. As described above, the proposed combination of Hedin in view of Gillick is not proper. For at least these reasons, the independent Claim 1 is allowable over of Hedin in view of Gillick.

The independent Claim 27 includes a method of recognizing spoken input received from a source of the spoken input. The method comprises receiving the spoken input from the source of the spoken input, performing a first pass speech recognition technique on the spoken input, forming first pass results, wherein the first pass results define one or more preliminary matching speech expressions, and selectively performing a second pass speech recognition technique on the spoken input according to the first pass results. As described above, the proposed combination of Hedin in view of Gillick is not proper. For at least these reasons, the independent Claim 27 is allowable over of Hedin in view of Gillick.

Claims 2-5, 7-9, 12, 17, 18, and 20 are each dependent upon the independent Claim 1. Claims 32-34, 37-39, and 45 are each dependent upon the independent Claim 27. As discussed above, Claims 1 and 27 are each allowable over the teachings of Hedin. Accordingly, Claims 2-5, 7-7, 12, 17, 18, 20, 32-34, 37-39, and 45 are each also allowable as being dependent upon allowable base claims.

Within the Office Action, the independent Claims 52 and 55 are allowed.

For the reasons given above, the Applicant respectfully submits that all of the claims are now in a condition for allowance, and allowance at an early date would be appreciated. Should the Examiner have any questions or comments, he is encouraged to call the undersigned at (408) 530-9700 to discuss the same so that any outstanding issues can be expeditiously resolved.

The Commissioner is authorized to charge any underpayment or credit any overpayment to Direct Deposit Account No. 18-1275 for any matter in connection with this response, including any fee for extension, which may be required.

Respectfully submitted,

HAVERSTOCK & OWENS LLP

Dated: 11-5-04

By:



Thomas B. Haverstock

Reg. No.: 32,571

Attorneys for Applicant

CERTIFICATE OF MAILING (37 CFR § 1.8(a))

I hereby certify that this paper (along with any referred to as being attached or enclosed) is being deposited with the U.S. Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to the: Commissioner for Patents, P.O. Box 1450 Alexandria, VA 22313-1450

HAVERSTOCK & OWENS LLP.

Date: 11/5/04

By:

